ISSN No:-2456-2165

# Agro-Genius: Crop Prediction using Machine Learning

Thayakaran Selvanayagam[1], Suganya S[2], Puvipavan Palendrarajah[3]
Mithun Paresith Manogarathash[4], Anjalie Gamage[5], Dharshana Kasthurirathna[6]
Faculty of Computing, Sri Lanka Institute of Information Technology (SLIIT)
Malabe, Sri Lanka.

**Abstract:- This paper present a way to aid farmers focusing on profitable vegetable cultivation in Sri Lanka. As agriculture creates an economic future for developing countries, the demand of modern technologies in this sector is higher. Key technologies used for this problem are Deep Learning, Machine Learning and Visualization. As the product, an android mobile application is developed. In this application the users should input their location to start the prediction process. Data preprocessing is started when the location is received to the system. The collected dataset divided into 3 parts. 80 percent for training, 10 percent for testing and 10 percent for validation. After that the model is created using LSTM RNN for vegetable prediction and ARIMA for price prediction. Finally, for given location profitable crop and predicted future price of vegetables are shown in the application. Other than the prediction, optimizing for multiple crop sowing according to the user requirements and visualizing cultivation and production data on map and graphs are also given in the application. This paper elaborates the procedure of model development, model training and model testing.**

***Keywords:-*** *Machine Learning, Android Application, Data preprocessing, LSTM, RNN, ARIMA, Linear Programming, Visualization, Polygons.*

## I. INTRODUCTION

A substantial percentage of the inhabitants of the country depend on the agriculture. The technological advancement in agriculture plays an important role in every farmer's life to earn good profit. But nowadays percentage of total GDP has been dropping. In 2005 the agriculture GDP share was 17.2% but in 2012 it has dropped to 11.1% and now it is even low [1]. Approximately 80% of the farmers are from rural areas so if crop production revenue goes down thus affect their lifestyle because of the industry level farms.

Apparently, Farmers' experience on the agriculture field involves in the crop prediction. Farmers who were in the rustic areas are cultivating according to their personal experience and knowledge due to absence of reliable and timely information. Since the modernization occupying the agriculture field rapidly by the introduction of superior seeds and different varieties and large number of crops which were cultivated by agricultural industries, the farmers are forced to adapt to this hasty change by cultivating more and more crops. Also, the main issue that small farmers are currently facing is that, they sow the crops according to their own experiences. But when they are cultivating and bringing them to market, they face difficulties to market their product at a reasonable price. It is because of large farms cultivating the same. As our country is small, products are distributed all over the country in between Districts (Dambulla to Jaffna, Dambulla to Petra, etc.). Because of this, small-scale rural farmers affected economically.

Nowadays weather condition is not like previous decades. Day by day it is changing because of the globalization, so farmers have faced difficulties to predict weather conditions. They may be some natural disaster which can also affects cultivation in a sudden. Without the weather, there are some major factors such as seasonal crop details, crop combination and suitable crop for given location which they must have knowledge of these things were gained from their past experience so without experience they can't get expected revenue. By considering these factors Agro-Genius system is recommended as a solution, hope that it will be very helpful for farmers to get expected revenue from their cultivation.

The main research problem is to help small medium farmers to increase revenue from their cultivation without getting affected by industrial level farmers and to reduce surplus marketing. Hitherto in our country there are no implemented techniques in usage, but agriculture department keeps so many raw data and using few in their website for public access, but it is not helpful to farmers. They cultivate according to their experience. When it's come to market, industry level farmers sell their product in a wholesale to all over the country at the same time rural farmers also bring their product, but they can't sell with a reasonable price. In this situation industry level farmers have no huge loss, but rural farmers loss their profits and even capital.

The principal scope of this research is; delivering a mobile application where all type of processing is done in the cloud-based system through the API calls. Which will be much helpful for the farmers and industries to select most profitable crop and its expected price during harvesting time. Further user can view the currently cultivated crop details in locations around the country on a map and user is able to optimize for profitable multiple crops for a specific land. The following data are collected from the relevant departments and from other third-party services.

ISSN No:-2456-2165

- Recommended crops details (location wise)
- Recommended crop harvesting duration.
- Past cultivation and production of vegetables.
- Currently cultivated crop which are updated biweekly.
- Weather forecast data.
- Combination of crops which will give more yield.
- Past market prices for each crop.

For the above-mentioned problem, Agro-Genius provide a solution using above listed past data. As it has more than 10 years of data it is not possible for human to predict from those huge amounts of data. So, to overcome this challenge Machine Learning would be more suitable technology now. In the above listed data, main data like past cultivation and production of vegetables, weather data and past market price report are timeseries data. To handle timeseries prediction some Machine Learning and Deep Learning algorithms are selected and explained in Methodology section below.

The research reduce problems of rural farmers and it suggest solution for more profitable cultivation. It helps farmers to take decisions when starting vegetable farms.

## II. RELATED WORK

In past years several systems have been proposed to implement crop prediction using machine learning techniques in several countries. Different Machine Learning algorithms were used for prediction. Multiple Leaner Regression (MLR) has been applied to predict on past data like year, area of sowing, rainfall, and yield and Data Mining methodology (Density – based clustering technique) is used to analyze and verify the result which was obtained from MLR [2]. On the other hand, for future forecasting was done by analyzing past historical price data, climate, location of market and planting area. Prediction was done for 15 market price data and 100 different crops using different algorithms like ARIMA, Artificial Neural Network (ANN), Response Surface Methodology (RSM) and calculate its Mean Absolute Percentage Error (MAPE). According to the lowest error percentage, many have selected ANN and PLS as prediction algorithms [3].

Arun Kumar et al… have proposed system to predict yield of the crop by analyzing past soil dataset, rainfall dataset, yield datasets. Prediction was done using K-Nearest Neighbor and Support Vector Machine algorithm and Least Squares algorithms [4]. Askunuri Manjula et al… has done crop prediction using weather forecasting, pesticides and fertilizers to be used and past revenue as input data. Multilinear Principal Component Analysis (MPCA) was used for feature reduction. Optimal Neural Network classifier (ONN) has been used for data prediction. Other than the prediction they consider preprocessing and feature reduction [5].

Leisa et al… have proposed Agriculture decision support framework for visualization and prediction of western Australian crop production system which will output visualizations of seasonal patterns of rainfall for individual district and show the effect of various scenarios. This system consists six major components which are data input, data mining, database, statistical analysis, prediction and visualization. Data input was done by Graphical User Interface (GUI). Data visualization done by two methods which are general trends and spatial interpolation. Data mining was done by the use of association rules which uses Apriori algorithms. [6]

There is lack of implemented systems used in other countries like United State. One is Field Check app which visualizes currently cultivated crop details in map, but it is visualizing some selected crops only [7]. Another one is Descartes Crop app which is forecasting crop yield for selected area in United States [8]. More over DEKALB is used to optimize multiple crop combination for small farms. Other than this implemented application Sri Lankan [9] and Taiwan [10] agriculture departments maintain some raw data, which are suitable crop for each land areas, past production for districts and historical price data. These help farmers, even though there are no prediction technology. Farmers have to go through the data one by one to make any decision.

In above mentioned proposed and implemented systems, they have not considered all the factors which are affecting farmers in the real world. If they consider main factors, then it will be more accurate. According to our country many factors affect farmers profit like weather, past cultivation and production details, market price etc. but there is no implemented system to guide Sri Lankan famers, so they failed to select profitable crop during seasons. In this system most of the features needed to solve the current problems are included and help farmers to select profitable crops. The application will provide predicted results such as most profitable crop and it's expected price according to the location and harvesting time. Also, users can view the cultivation details visualized in map as it will be more effective than statistical data.

## III. METHODOLOGY

This proposed system contains four main components such as crop prediction, price prediction, visualization and optimization. Each component uses different Machine Learning algorithms and techniques, they are Long-Short Term Memory (LSTM), Auto Regressive Integrated Moving Average (ARIMA), Linear Programming and Gastner-Newman Cartogram techniques to predict and visualize raw datasets.

*A. Long-Short Term Memory –*

LSTM is type of neural network which perform better result in time series prediction. Purpose of this algorithm is avoiding the long-term dependency and LSTM is called cell state which contain different gate. [11]

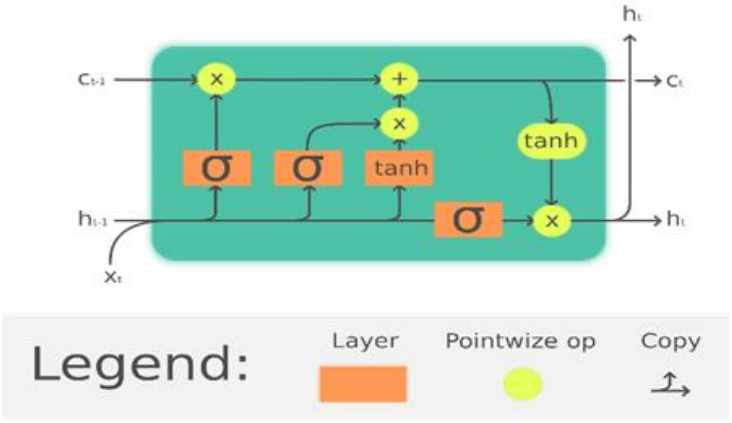


Fig 1:- LSTM cell state diagram, Image downloaded from https://commons.wikimedia.org/wiki/File:The_LSTM_cell.png#/media/File:The_LSTM_cell.png

The above diagram shows different notations are used, where $X_t$ denote input vector, $H_{t-1}$ denote Previous cell output, $C_{t-1}$ denote previous cell memory in addition $H_t$ is Current cell output and $C_t$ denote Current cell Memory. Following formulas are used to find the values of above-mentioned notations.

$F_t = \sigma (X_t * U_f + H_{t-1} * W_f)$
$C' = \tanh (X_t * U_c + H_{t-1} * W_c)$
$I_t = \sigma (X_t * U_i + H_{t-1} * W_i)$
$O_t = (X_t * U_o + H_{t-1} * W_o)$
$C_t = f_t * C_{t-1} + I_t * C'_t$
$H_t = O_t * \tanh (C_t)$

*B. Auto Regressive Integrated Moving Average (ARIMA) –*

ARIMA is statistical analysis model that is used for time series data prediction. ARIMA is divided into 3 components such as Autoregression (AR), Integrated (I) and Moving Average (MA). ARIMA model is classified as ARIMA(p,d,q) where p denotes the number of lag observation in the model, d denotes the number of times that the raw observations are differenced and q denotes the moving average window size[12]. Following formula is used to find the price forecasting where μ denote constant value.

$\hat{Y}_t = \mu + Ø_1 y_{t-1} + .. + Ø_p y_{t-p} - Θ_1 e_{t-1} .. - Θ_q e_{t-q}$

*C. Linear Programming*

It is an optimization technique, by using this can find the optimum points of object function.

*D. Gastner-Newman Cartogram*

It is a technique for representing data for locations. Cartogram is a powerful approach to map data [13]. It provides strong visual for numerical area also this technique doesn't need data to be normalized. Comparing other technique, this is easy to visualize each polygon.

➢ *Used Datasets*

This system (Agro-Genius) is fully based on statistical data and most of the data are from Agriculture Department of Sri Lanka. There were data collected for more than 10 years with different seasons in Sri Lanka like Yala and Maha. Below are the important factors that affect agricultural crop yield. which were selected for this research.

1. *Crop production and extent:* crop cultivated area in hectares and total production in metric ton for every year in each district in Sri Lanka for two main season such as Yala and Maha.
2. *Recommended crop:* each district located with different soil type therefore each district has recommended crop according to the soil type.
3. *Crop duration:* number of days that take to harvest from seeds for search type of crop.
4. *Crop combination:* crop type which can be sewed together in same land.
5. *Current cultivation extent:* biweekly updated data of cultivation extent of crops in the present time.
6. *Price reports:* Crop market price in main markets in selected districts were taken from Central Bank of Sri Lanka.
7. *Weather data:* weather data for coming weeks are received from AccuWeather.

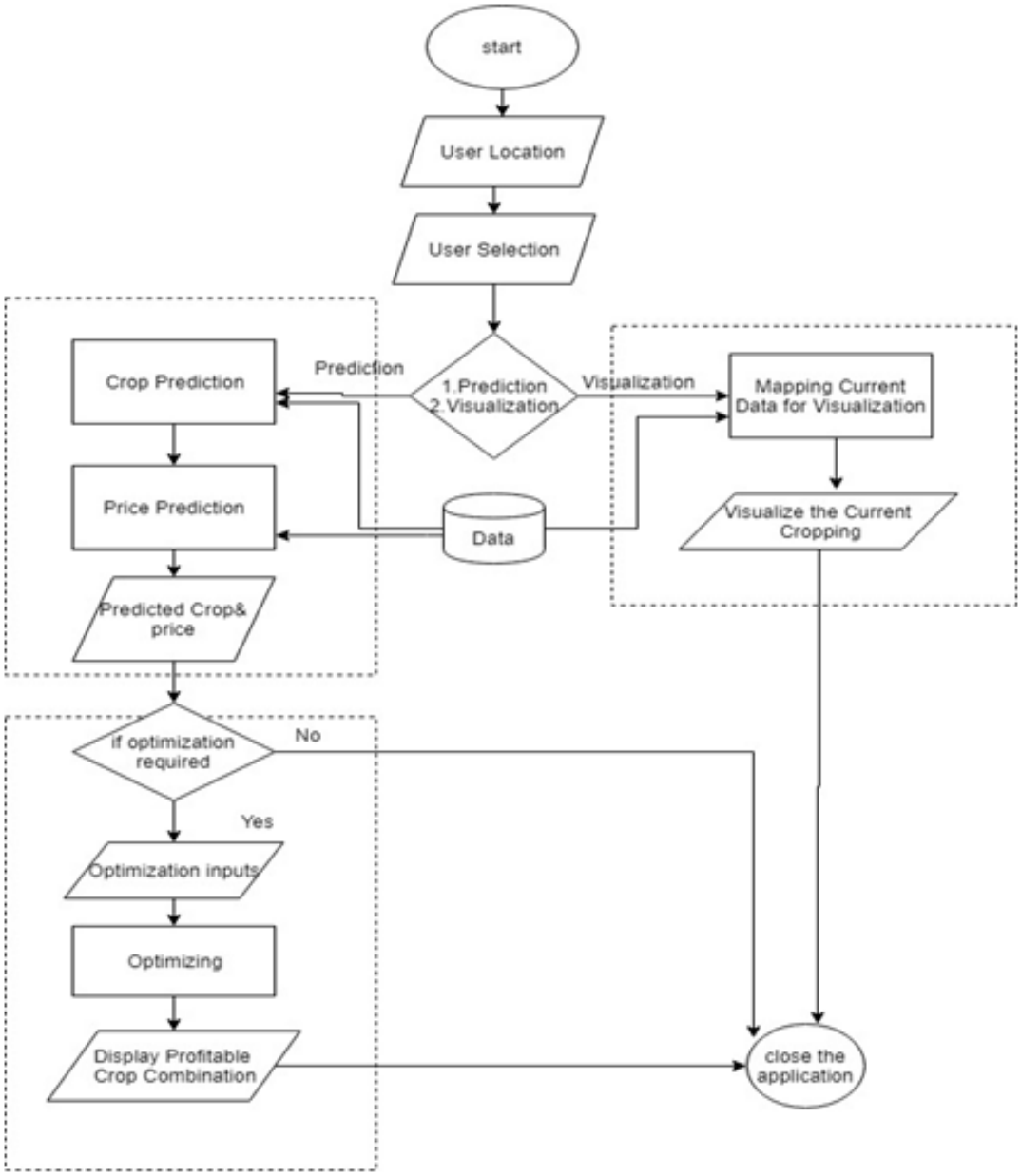


Fig 2:- flow of the system

The following figure shows the workflow of the system.

Location of user will be inputted to the system. Otherwise user should manually input the location where he/she wants to get the prediction. In the prediction component, with the given location the existing trained model will be analyzed, and it will predict suitable crops and then the predicted crop will be analyzed with the price prediction model and expected price will be listed. And then if the user wants to optimize the crops that were predicted for a better profitable combination user can proceed to the optimization component. For more detailed explanation of current cropping around the country, the current cropping data are mapped in to the map. Another view can be obtained to visualize the past cropping pattern.

*i. Crop & Price Prediction*

To maximize crop profit, appropriate crop selection will play a vital role. In this paper profitable crop selection based on statistical data like past production data, recommended crop details for each district, past price data and weather forecasting data are used. To analyze these data RNN & LSTM technique was used. After crop selection, for those selected crops expected price in harvesting time will be predict using ARIMA technique.

*ii. Visualization*

Currently cultivated crop details are displayed in Sri Lankan map with different colors using Gatogram technique & past cultivation details will be visualized using a bar chart.

*iii. Multi Crop Optimization*

Optimize for profitable crop combinations using linear programming for the predicted crops with optimization inputs like suitable crop combination.

## IV. RESULTS & DISCUSSIONS

As mentioned, above this system contains four component all component depends on each other because output of the one component will be the input of another component so the accuracy of each component will affect the entire system, And this is fully based on dataset prediction, so accuracy is very important, Since this system is fully based on dataset all collected raw datasets are normalized according to the model requirement within that normalized data 80% of the data was taken as training data, 10% taken for validating and 10% taken for testing.

*A. Crop Prediction*

This prediction involves many datasets. All datasets are preprocessed according to the user location and trained using LSTM & Random Forest Regression model. For the comparison, the districts where the user's market is considered. Here past data set of production for last 10 years for each district were used. LSTM gives more accuracy for this time series data still current data amount is not enough for LSTM to give higher accuracy as it has only seasonal cultivation and production. So, Random forest is working better for this dataset.

| Year | OD_Area_Hect | OD_Prod_MT | CD_Area_Hect | CD_Prod_MT |
|---|---|---|---|---|
| 2009 | 196.666667 | 1525.333333 | 161.0 | 1441.0 |
| 2010 | 166.000000 | 1353.666667 | 222.0 | 1684.0 |
| 2011 | 153.333333 | 1390.666667 | 162.0 | 1244.0 |
| 2012 | 175.333333 | 1533.000000 | 167.0 | 1153.0 |

Fig 3:- Normalized cultivation & production data

| Year | Season | CD_Avg_Price | Matale | Kandy | Nuwaraeliya | OD_Avg_Price |
|---|---|---|---|---|---|---|
| 2008 | Maha | 86.527192 | 77.485479 | 119.964110 | 53.677055 | 83.708881 |
| | Yala | 87.897850 | 60.921215 | 98.059813 | 57.897664 | 72.292897 |
| 2009 | Maha | 103.669247 | 122.417055 | 98.943699 | 81.681644 | 101.014132 |
| | Yala | 108.821698 | 110.649623 | 129.722358 | 79.172075 | 106.514686 |
| 2010 | Maha | 106.410068 | 307.179658 | 46.601781 | 43.242466 | 132.341301 |
| | Yala | 107.777358 | 306.534528 | 45.719151 | 39.649245 | 130.634308 |
| 2011 | Maha | 99.587603 | 163.448836 | 11.086918 | 40.233425 | 71.589726 |

Fig 4:- Normalized price data

*B. Crop Price Prediction*

Price data for last 10 years were used to predict expected price for each crop. This price differs from each district, so the location was considered as well. Price data filtered by location and those filtered data were processed using ARIMA and LSTM models and Mean Absolute Percentage Error was calculated. Since these price datasets is time series and huge so ARIMA is the best selection for price prediction and accuracy also higher than LSTM. Result of each algorithm are shown below.

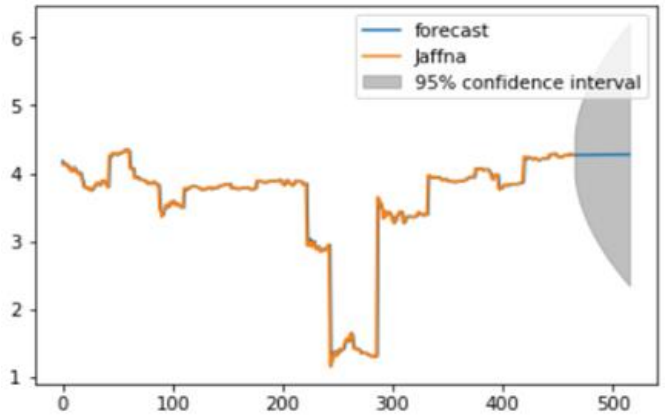


Fig 5:- Predicted price forecasting using ARIMA model

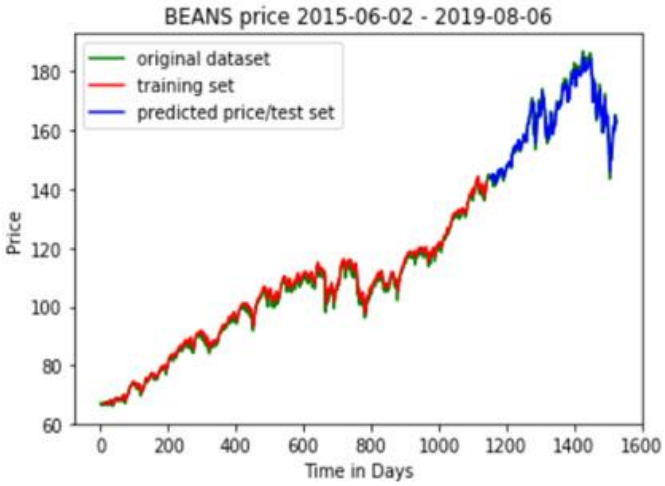


Fig 6:- Predicted price forecasting using LSTM model

*C. Visualization*

Currently cultivated data were analyzed and visualized in the Sri Lankan map using Cartogram technique. Which locate the cultivated geographical location as per the latitude and longitude value. Mainly 6 districts were considered in Sri Lanka (Matale, Kandy, Nuwara Eliya, Jaffna, Kilinochchi, Mullaitivu) Within those districts cultivated areas and it's details like cultivated area in hectare and harvesting time were taken. This visualized map will be updated according to the harvesting time.

ISSN No:-2456-2165

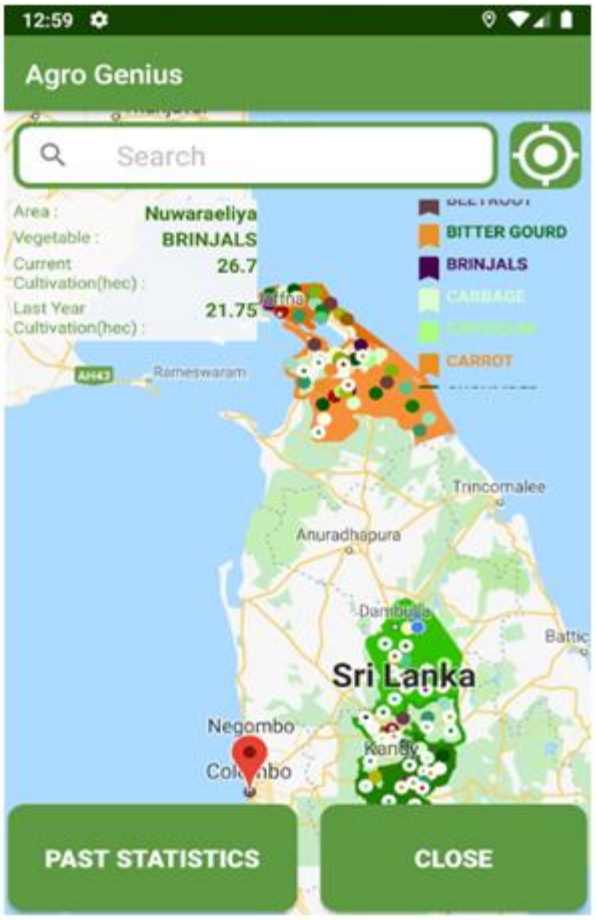


Fig 7:- Visualized currently cultivated data

Also, past statistical results analyzed would be visualized using bar graph that indicate previously cultivated details in hectares according to the district for each crop.

*D. Multi Crop Optimization*

Linear programming model was used with input parameters such as predicted crop list, crop suitability, area in hectare and no of crops that user want to cultivate. Which give 89.66% accuracy for prediction with the available data set.

## V. CONCLUSION & FUTURE WORK

Agriculture is the major economic force in Sri Lanka. It has moderate climate throughout the year in most parts of the country [14]. As the country is small, cultivated crops are distributed all over the country, because of that a reasonable market price is remaining as a challenging issue for farmers. To overcome this problem, Agro-genius application advice to predict the most profitable crops and its expected price during harvesting time according to the location, by predicting different historical raw datasets using different machine learning algorithms like LSTM & RNN, ARIMA, Linear Programming (LP), Gastner Newman Cartogram algorithm etc. Also, currently cultivated crop details are visualized in a map with details, which will help farmers to view the nearby district cultivation details. This system helps farmers to take correct decision for selecting suitable crops, which will maximize the profit.

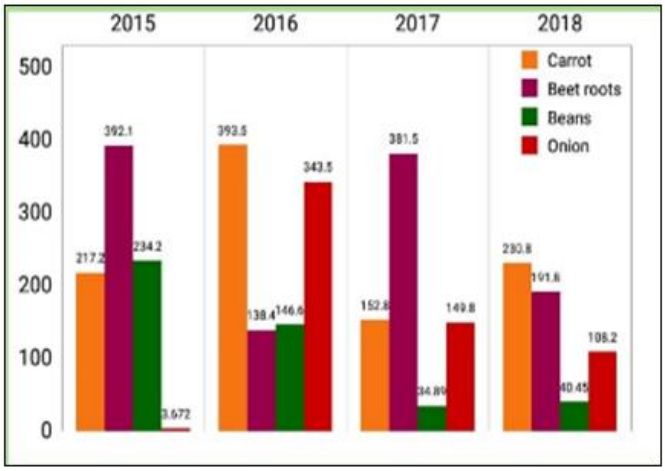


Fig 8:- Past production details

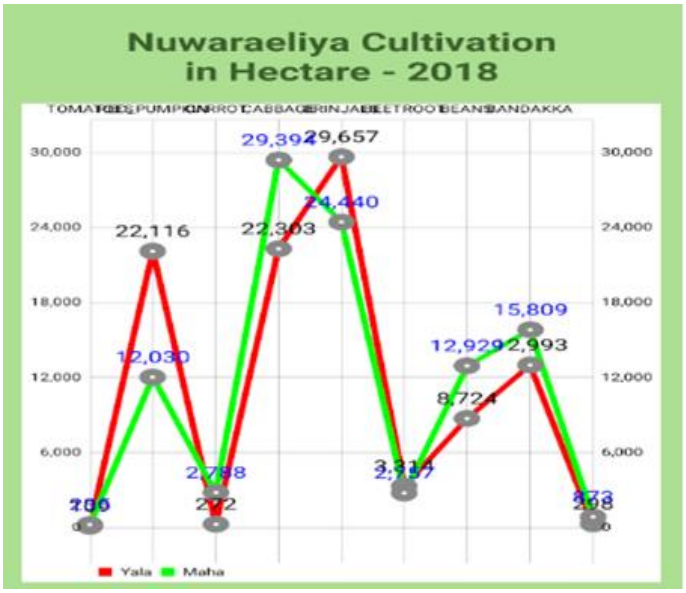


Fig 9:- Past cultivation details

This system considers main features which impact the profit by taking data in six districts. In future the system can be expanded by considering more features like soil type and water level and so on. Also expand to provide fertilization calendar and guidelines which will helps farmers who have no experience about crops. Also, in the system can be modified to receive data from IoT devices without depending on raw data. Other than that, this system can be developed for other platforms as well.

## REFERENCES

[1]. Ft.lk. (2019). Sri Lanka heading for an agricultural issue in 2017? [Online] Available at: http://www.ft.lk/columns/sri-lanka-heading-for-an-agriculturalissue-in-2017/4-588715 [Accessed 8 Mar. 2019].
[2]. D Ramesh and B Vishnu Vardhan, "ANALYSIS OF CROP YIELD PREDICTION USING DATA MINING TECHNIQUES", International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 23217308.
[3]. Yung-Hsing Peng, Chin-Shun Hsu and Po-Chuang Huang, "Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets", 978-1-46739606-6/15/$3l.00 ©2015 IEEE.
[4]. Arun Kumar, Naveen Kumar and Vishal Vats , "EFFICIENT CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHMS", International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 06 | June-2018 e-ISSN: 2395-0056 |p-ISSN: 2395-0072.
[5]. Aakunuri Manjula and Dr. G.Narsimha, "Crop Yield Prediction with Aid of Optimal Neural Network in Spatial Data Mining: New Approaches", International Journal of Information & Computation Technology. ISSN 09742239 Volume 6, Number 1 (2016), pp. 25-33.
[6]. Leisa J. Armstrong and Sreedhar A. Nallan, "Agricultural Decision Support framework for visualization and prediction of western Australian crop production", - IEEE 978-9-3805-4421-2/16 – 2016.
[7]. Play.google.com. (2019). [online] Available at: https://play.google.com/store/apps/details?id=com.field watch.FieldCheck&hl=en _US [Accessed 8 Mar. 2019].
[8]. Anon, (2019). [online]Available at: https://itunes.apple.com/us/app/%20escartes-crops/id1140422866?ls=1&mt=8 [Accessed 8 Mar. 2019].
[9]. User, S. (2019). Department Of Agriculture. [online] Doa.gov.lk. Available at: https://www.doa.gov.lk/en/ [Accessed 8 Mar. 2019].
[10]. Amis.afa.gov.tw. (2019). Agricultural wholesale market trading market station. [online] Available at: http://amis.afa.gov.tw/main/About.aspx [Accessed 8 Mar. 2019].
[11]. Colah.github.io. (2019). Understanding LSTM Networks -- colah's blog. [online] Available at: https://colah.github.io/posts/2015-08-Understanding-LSTMs/ [Accessed 6 Aug. 2019].
[12]. Investopedia. (2019). Autoregressive Integrated Moving Average (ARIMA). [online] Available at: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp [Accessed 6 Aug. 2019].
[13]. Arcgis.com. (2019). [online] Available at: https://www.arcgis.com/home/item.html?id=4b8c9ce99a5749e298bb96366692f35d [Accessed 6 Aug. 2019].
[14]. User, S. (2019). Ministry of Agriculture - Sri Lanka - Overview. [Online] Agrimin.gov.lk. Available at: http://www.agrimin.gov.lk/web/index.php/aboutus/overview123 [Accessed 8 Mar. 2019].